Springer Link

- Original Paper
- Published: 11 June 2021

Machine learning prediction of breast cancer survival using age, sex, length of stay, mode of diagnosis and location of cancer

- Hilary I. Okagbue,
- Patience I. Adamu,
- Pelumi E. Oguntunde,
- Emmanuela C. M. Obasi &
- Oluwole A. Odetunmibi

Health and Technology **volume 11**, pages887–893 (2021)Cite this article

Abstract

Breast cancer is one of the leading causes of death in females and survival depends on early diagnosis and treatment. This paper applied machine learning techniques in prediction of breast cancer survival (dead or alive) using age, sex, length of stay, mode of diagnosis and location of cancer as predictors (independent variables). The data was obtained from the outpatient department of the University of Ilorin Teaching Hospital, Ilorin, Nigeria. The sample size of 300 consists of 175 females and 25 males who were admitted at the hospital and treated for breast cancer. The patients were later discharged or died. Adaptive boosting (AdaBoost) performed best out of the data mining models used in the classification in all the three cases where the target class is average over classes, alive or dead. The AdaBoost performed best with the classification accuracy and area under curve (AUC) of 98.3% and 99.9% respectively. Furthermore, a probe on the prediction by AdaBoost showed that

the probability of dead due to breast cancer is 0.47, which the length of stay hugely contributed to the high probability, location of breast cancer and mode of diagnosis contributed minimally while age and sex contributed insignificantly. The high probability of breast cancer mortality predicted in this paper is a call for concern as early detection of breast cancer, routine breast examination and breast cancer awareness are crucial in increasing the probability of survival. The results can be used to design a decision support system that can increase the chances of breast cancer survival.

References

1. 1.

Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. CA Cancer J Clin. 2011;61(2):69–90. https://doi.org/10.3322/caac.20107.

Article **Google Scholar**

2. 2.

World Health Organization fact sheets for breast cancer. 2021. https://www.who.int/news-room/fact-sheets/detail/breast-cancer. Accessed 3 Jun 2021.

3. 3.

Gucalp A, Traina TA, Eisner JR, Parker JS, Selitsky SR, Park BH, Elias AD, Baskin-Bey ES, Cardoso F. Male breast cancer: a disease distinct from female breast cancer. Breast Cancer Res Teat. 2019;173(1):37–48. https://doi.org/10.1007/s10549-018-4921-9.

Article **Google Scholar**

4. 4.

Adamu PI, Adamu MO, Okagbue HI, Opoola L, Bishop SA. Survival Analysis of Cancer Patients in North Eastern Nigeria from 2004–2017 – A Kaplan - Meier Method. Open Access Maced J Med Sci. 2019;7(4):642–9. https://doi.org/10.3889/OAMJMS.2019.109.

Article **Google Scholar**

5. 5.

Adamu PI, Oguntunde PE, Okagbue HI, Agboola OO. Statistical data analysis o5 cancer incidences in insurgency affected states in Nigeria. Data Brief. 2018;18:2029–46. https://doi.org/10.1016/j.dib.2018.04.135.

Article **Google Scholar**

6. 6.

Beddoe AM. Elimination of cervical cancer: challenges for developing countries. ecancermedicalscience. 2019;13:97. https://doi.org/10.3332/ecancer.2019.975.

7. 7.

Sassi F. Global public health challenges, fiscal policies, and yellow vest. Lancet. 2019;393(10173):745–6. https://doi.org/10.1016/S0140-6736(19)30253-3\.

Article **Google Scholar**

8. 8.

Chougrad H, Zouaki H, Alheyane O. Multi-label transfer learning for the early diagnosis of breast cancer. Neurocomp. 2020;392:168–80.

Article **Google Scholar**

9. 9.

Sung H, Siegel RL, Torre LA, Pearson-Stuttard J, Islami F, Fedewa SA, Goding Sauer A, Shuval K, Gapstur SM, Jacobs EJ, Giovannucci EL, Jemal A. Global patterns in excess body weight and the associated cancer burden. CA Cancer J Clin. 2019;69(2):88–112.

Google Scholar

10.10.

Jansen T, Geleijnse G, van Maaren M, Hendriks MP, Ten Teije A, Moncada-Torres A. Machine learning explainability in breast cancer survival. Stud Heal Tech Inform. 2020;270:307–11.

Google Scholar

11.11.

Hussain SS, Gandhi O, Sajida SS. Breast cancer detection and classification using NSST-based modified FPCM with ANN. Int J Adv Sci Tech. 2020;29(8):1567–76.

Google Scholar

12.12.

Dawngliani MS, Chandrasekaran N, Lalmuanawma S, Thangkhanhau H. Prediction of Breast Cancer Recurrence Using Ensemble Machine Learning Classifiers. Adv Intel Syst Comput. 2020;1145:232–44.

Google Scholar

13.13.

Kwon H, Park J, Lee Y. Stacking ensemble technique for classifying breast cancer. Healthc Inform Res. 2019;25(4):283–8.

Article **Google Scholar**

14.14.

Nallamala SH, Mishra P, Koneru SV. Qualitative metrics on breast cancer diagnosis with neuro fuzzy inference systems. Int J Adv Trends Comp Sci Engine. 2019;8(2):259–64.

Article **Google Scholar**

15.15.

Ming CO, Viassolo V, Probst-Hensch N, Chappuis PO, Dinov ID, Katapodi MC. Machine learning techniques for personalized breast cancer risk prediction: Comparison with the BCRAT and BOADICEA models. Breast Cancer Res. 2019;21(1):75.

Article **Google Scholar**

16.16.

Rajamohana SP, Umamaheswari K, Karunya K, Deepika R. Analysis of Classification Algorithms for Breast Cancer Prediction. Adv Intel Syst Comp. 2020;1042:517–28.

Article **Google Scholar**

17.17.

Kleinlein R, Riaño D. Persistence of data-driven knowledge to predict breast cancer survival. Int J Med Inform. 2019;129:303–11.

Article **Google Scholar**

18.18.

Akinsola JET, Adeagbo MA, Awoseyi AA. Breast cancer predictive analytics using supervised machine learning techniques. Int J Adv Trends Comp Sci Engine. 2019;8(6):3095–104.

Article **Google Scholar**

19.19.

Malik N, Bharat VB, Tiwari SP, Singla J. Study of detection of various types of cancers by using deep learning: A survey. Int J Adv Trends Comp Sci Engine. 2019;8(4):1228–33.

Article **Google Scholar**

20.20.

Coccia M. Deep learning technology for improving cancer care in society: New directions in cancer imaging driven by artificial intelligence. Tech Society, 2020;60:101198.

21.21.

Alzahrani MY, Alahmadi AH. Breast cancer image classification using the convolution neural network. Int J Adv Trends Comp Sci Engine. 2019;8(6):3447–53.

Article **Google Scholar**

22.22.

Kumar A, Singh SK, Saxena S, Lakshmanan K, Sangaiah AK, Chauhan H, Shrivastava S, Singh RK. Deep feature learning for histopathological image classification of canine mammary tumors and human breast cancer. Info Sci. 2020;508:405–21.

Article **Google Scholar**

23.23.

Piantadosi G, Sansone M, Fusco R, Sansone C. Multi-planar 3D breast segmentation in MRI via deep convolutional neural networks. Artif Intel Med. 2020;103:101781.

24.24.

Owoyemi A, Owoyemi J, Osiyemi A, Boyd A. Artificial intelligence for healthcare in Africa. Front Digital Heal. 2020;2:6. https://doi.org/10.3389/fdgth.2020.00006.

Article **Google Scholar**

25.25.

Jedy-Agba E, McCormack V, Adebamowo C, dos-Santos-Silva I. Stage at diagnosis of breast cancer in sub-Saharan Africa: a systematic review and meta-analysis. Lancet Glob Heal. 2016;4(12):e923-e935. https://doi.org/10.1016/S2214-109X(16)30259-5.

26.26.

Wahl B, Cossy-Gantner A, Germann S, Schwalbe NR. Artificial intelligence (AI) and global health: how can AI contribute to health in resource-poor settings? BMJ Glob Heal. 2018;3: e000798. https://doi.org/10.1136/bmjgh-2018-000798.

Article **Google Scholar**

27.27.

Ukwenya AY, Yusufu LMD, Nmadu PT, Garba ES, Ahmed A. Delayed treatment of symptomatic breast cancer: the experience from Kaduna. Nigeria South Afr J Surgery. 2008;46(4):106–10.

Google Scholar

28.28.

Gwagwa A, Kraemer-Mbula E, Rizk N, Rutenberg I, De Beer J. Artificial intelligence (AI) deployments in Africa: benefits, challenges and policy dimensions. The African Journal of Information and Communication. 2020;2020(26):1–28.

Article **Google Scholar**

29.29.

Idri A, Bouchra EO, Hosni M, Abnane I. Assessing the impact of parameters tuning in ensemble based breast Cancer classification. Heal Tech. 2020;10(5):1239–12551. https://doi.org/10.1007/s12553-020-00453-2.

Article **Google Scholar**

30.30.

Hasan N, Bao Y. Understanding current states of machine learning approaches in medical informatics: a systematic literature review. Heal Tech. 2021. https://doi.org/10.1007/s12553-021-00538-6.

Article **Google Scholar**

31.31.

Tsarouchi MI, Vlachopoulos GF, Karahaliou AN, Costaridou LI. Diagnostic value of apparent diffusion coefficient lesion texture biomarkers in breast MRI. Heal Tech. 2020;10(4):969–78. https://doi.org/10.1007/s12553-020-00452-3.

Article **Google Scholar**

32.32.

Vilardell M, Buxó M, Clèries R, Martínez JM, Garcia G, Ameijide A, Font R, Civit S. Missing data imputation and synthetic data simulation through modeling graphical probabilistic dependencies between variables (ModGraProDep): An application to breast cancer survival. Artif Intel Med. 2020;107:101875.

33.33.

Sinha A, Sahoo B, Rautaray SS, Pandey M. Analysis of Breast Cancer Dataset Using Big Data Algorithms for Accuracy of Diseases Prediction. Lect Notes Data Engine Comm Technol. 2020;44:271–7.

Article **Google Scholar**

34.34.

Goyal K, Aggarwal P, Kumar M. Prediction of Breast Cancer Recurrence: A Machine Learning Approach. Adv Intel Syst Comp. 2020;990:101–13.

Article **Google Scholar**

35.35.

Juneja K, Rana C. An improved weighted decision tree approach for breast cancer prediction. Int J Info Tech. 2020;12(3):797–804.

Google Scholar

36.36.

Mishra AK, Roy P, Bandyopadhyay S. Binary Particle Swarm Optimization Based Feature Selection (BPSO-FS) for Improving Breast Cancer Prediction. Adv Intel Syst Comp. 2021;1164:373–84.

Article **Google Scholar**

37.37.

Lo Gullo R, Eskreis-Winkler S, Morris EA, Pinker K. Machine learning with multiparametric magnetic resonance imaging of the breast for early prediction of response to neoadjuvant chemotherapy. Breast. 2020;49:115–22.

Article **Google Scholar**

38.38.

Sugasawa S, Noma H. Estimating individual treatment effects by gradient boosting trees. Stat Med. 2019;38(26):5146–59.

MathSciNet **Article** **Google Scholar**

39.39.

Kakudji BK, Mwila PK, Burger JR, Du Plessis JM. Epidemiological, clinical and diagnostic profile of breast cancer patients treated at Potchefstroom regional hospital, South Africa, 2012–2018: An open-cohort study. Pan Afr Med J. 2020;36:9. https://doi.org/10.11604/pamj.2020.36.9.21180.

40.40.

Olaogun JG, Omotayo JA, Ige JT, Omonisi AE, Akute OO, Aduayi OS. Socio-demographic, pattern of presentation and management outcome

of breast cancer in a semi-urban tertiary health institution. Pan Afr Med J. 2020;36:363. https://doi.org/10.11604/pamj.2020.36.363.17866.

41.41.

Howard FM, Olopade OI. Epidemiology of Triple-Negative Breast Cancer: A Review. Cancer J. 2021;27(1):8–16.

Article **Google Scholar**

42.42.

Arasi MA, Babu S. Survey of machine learning techniques in medical imaging. International J Adv Trends Comp Sci Engine. 2019;8(5):210–6.

Google Scholar

43.43.

Nourelahi M, Zamani A, Talei A, Tahmasebi S. A model to predict breast cancer survivability using logistic regression. Middle East J Cancer. 2019;10(2):132–8.

Google Scholar

44.44.

Ganggayah MD, Taib NA, Har YC, Lio P, Dhillon SK. Predicting factors for survival of breast cancer patients using machine learning techniques. BMC Med Inform Dec Making. 2019;19(1):48.

Article **Google Scholar**

45.45.

Mihaylov I, Kańduła M, Krachunov M, Vassilev D. A novel framework for horizontal and vertical data integration in cancer studies with application to survival time prediction models. Biol Direct. 2019;14(1):22.

Article **Google Scholar**

46.46.

Simsek S, Kursuncu U, Kibis E, AnisAbdellatif M, Dag A. A hybrid data mining approach for identifying the temporal effects of variables associated with breast cancer survival. Expert Syst Appl. 2020;139:112863.

47.47.

Boeri C, Chiappa C, Galli F, De Berardinis V, Bardelli L, Carcano G, Rovera F. Machine Learning techniques in breast cancer prognosis prediction: A primary evaluation. Cancer Med. 2020. https://doi.org/10.1002/cam4.2811.

Article **Google Scholar**

48.48.

Oguntunde PE, Adejumo AO, Okagbue HI. Breast cancer patients in Nigeria: Data exploration approach. Data Brief. 2017;15:47–57.

Article **Google Scholar**

49.49.

Mishra AK, Roy P, Bandyopadhyay S, Das SK. Breast ultrasound tumour classification: A Machine Learning—Radiomics based approach. Expert Syst. 2021. https://doi.org/10.1111/exsy.12713.

Article **Google Scholar**

50.50.

Roy SD, Das S, Kar D, Schwenker F, Sarkar R. (2021). Computer aided breast cancer detection using ensembling of texture and statistical image features. Sensors. 2021;21(1):3628. https://doi.org/10.3390/s21113628.

51.51.

Sedighi-Maman Z, Mondello A. A two-stage modeling approach for breast cancer survivability prediction. Int J Med Inform. 2021;2021(149): 104438. https://doi.org/10.1016/j.ijmedinf.2021.104438.

Article **Google Scholar**

52.52.

Song J, Zhang Y, Wang S, Liu Z, Sun D. Neural network combining X-ray and ultrasound in breast examination. Neural Comp Appl. 2021. https://doi.org/10.1007/s00521-021-05882-2.

Article **Google Scholar**

53.53.

Massafra R, Latorre A, Fanizzi A, Bellotti R, Didonna V, Giotta F, La Forgia D, Nardone A, Pastena M, Ressa CM, Rinaldi L, Russo AOM. A Clinical Decision Support System for Predicting Invasive Breast Cancer Recurrence: Preliminary Results. Front Oncol. 2021;11: 576007. https://doi.org/10.3389/fonc.2021.576007.

Article **Google Scholar**

54.54.

Siddiqui SY, Naseer I, Khan MA, Mushtaq MF, Naqvi RA, Hussain D, Haider A. Intelligent breast cancer prediction empowered with fusion and deep learning. Comp Mat Cont. 2021;67(1):1033–1049. https://doi.org/10.32604/cmc.2021.013952.

55.55.

Olow AK, Veer L, Wolf DM. Toward developing a metastatic breast cancer treatment strategy that incorporates history of response to previous treatments. BMC Cancer. 2021;21(1):212. https://doi.org/10.1186/s12885-021-07912-7.

Article **Google Scholar**

Download references

Acknowledgements

The authors appreciate the efforts of the anonymous reviewers toward this publication. The financial support from Covenant University, Nigeria is also deeply appreciated.

Funding

Covenant University.

Author information

Affiliations

1. **Department of Mathematics, Covenant University, Ota, Nigeria**
   Hilary I. Okagbue, Patience I. Adamu, Pelumi E. Oguntunde & Oluwole A. Odetunmibi

2. **Department of Computer Science and Informatics, Federal University, Otuoke, Nigeria**
Emmanuela C. M. Obasi

Contributions

All the authors contributed equally in the research that birthed the manuscript.

Corresponding author

Correspondence to Hilary I. Okagbue.

Ethics declarations

Research involving human participants

Not applicable. No experiment was performed on animal or human subject (s).

Informed consent

The details are available on https://www.sciencedirect.com/science/article/pii/S2352340917304158

Conflict of interest

None declared.

Confidentiality

Not applicable. The data paper from where the data was obtained was duly acknowledged and cited.

Additional information

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Rights and permissions

About this article

Cite this article

Okagbue, H.I., Adamu, P.I., Oguntunde, P.E. *et al.* Machine learning prediction of breast cancer survival using age, sex, length of stay, mode of diagnosis and location of cancer. *Health Technol.* **11,** 887–893 (2021). https://doi.org/10.1007/s12553-021-00572-4

Download citation

- Received28 March 2021
- Accepted09 June 2021
- Published11 June 2021
- Issue DateJuly 2021
- DOIhttps://doi.org/10.1007/s12553-021-00572-4

Keywords

- **Breast cancer**
- **Data mining**
- **Machine learning techniques**
- **Nigeria**
- **Survival**